調査の誤差

浅井 晃*

ON SURVEY ERRORS

Akira Asai*

1. 調査誤差を認識せよ

主成分分析・重回帰分析・判別関数・因子分析・数量化法・クラスター分析、多次元尺度構成法等々と、行動科学のデータ解析に利用できる手法が数多く開発され、一方 SPSS, BMDP をはじめとする多種類のプログラム・パッケージが用意され、データ解析がきわめて多面的かつ容易になってきたことは誠に喜ばしいことである。ところが、行動科学のデータ解析にこれらの手法を利用した数多くの結果を見ていると、かなりのものに気に掛かることがある。それは、調査誤差の問題である。ここでいう調査誤差とは、後述するように、広義のものである。多変量統計解析の手法には誤差処理がなされている、という反論もあるかもしれないが、それらはデータ内の誤差変動の処理に過ぎないといっても過言ではない。そこで、新しいことではないが、再認識を求める意味で（広義の）誤差に関連した問題をいくつか、以下で論じよう。

ここでいう広義の誤差とは次のものを総称することにする。

- ① 目的とする母集団からのランダム標本と考え難いデータにもとづく結果の持つ誤差。
- ② 再現性、普遍性の乏しいデータ解析の持つ誤差。
- ③ ランダム標本にもとづく誤差、すなわち標本誤差。
- ④ 調査対象の偽り・調査員のミス・その他データ収集の過程で導入される誤差、すなわち非標本誤差。

まず①について、数理統計学の理論のほとんどは、母集団からの単純ランダム標本を前提としている。したがって、集められたデータに統計理論を適用するには、標本の単純ランダム性を主張できるじゅうぶんな根拠がな

ければならない。もし目標とする母集団に対して、標本の単純ランダム性が保証できる対応集団が限定されたものならば、目標母集団を修正して対応集団に改めなければならない。例えば、政治意識に関するアンケート調査を街頭で行って得たデータを分析した結果は、一般有権者を対応母集団とするものとは考えられないし、いさかもそうであるかの誤解を与えるような公表は許されない。自然現象を測定して得たデータの場合には、例外も多いが、単純ランダム性をたしかめるための特別の検討を必要としないことも多いが、社会現象・心理現象など人間の行動に関連したデータを集める場合には、母集団と標本との関係をじゅうぶん吟味しなければならない。むしろ、計画的にランダムサンプリングを実行すべきである。標本調査法とかランダムサンプリングの理論といわれるものは、主としてこの目的のために作られたものである。官庁統計調査・世論調査・市場調査のほか行動科学関係のアンケート調査などで広く適用され、あるいは適用すべく努力がなされているが、一方その報告から判断して、標本調査法を適用するよう強く忠告したいものも多い。調査対象の選定が安易に行われたデータの分析結果はケース・スタディに過ぎないともいえる。多変量統計解析法・数量化法・因子分析法などの理論と同一水準で標本調査法も重視すべきである。費用・労力等の制約から、目標母集団からのランダムサンプリングが実行できない場合は、ランダムサンプリングが可能な母集団に目標を改定すべきである。

次に②について、得られたデータが母集団のすべてである特殊な場合を除けば、得られたデータは、上述のように、目標母集団からのランダムサンプリングの所産と考えるべきである。そして、そのデータを分析することは、そのランダム性を除去して普遍的な法則性を導くことでなければならない。データに対して数学的方法を駆使して法則性を導いたとしても、母集団からのランダム

* 千葉大学理学部数学科 (Department of Mathematics, Faculty of Science, Chiba University)

サンプリングをやり直して得たデータから同じ法則性が得られるかどうかは疑問である。比喩的に、データにもとづく平面上の n 個の点を通る曲線として、数学的に n 次多項式を決めることができるが、データを取り直せば多分同じ多項式にはならないことは確実で、したがってこのような曲線を求めることは無意味ともいえる。すなわち、分析結果が再現性・普遍性を持つものであってはじめてその分析が有効であったといえる。もちろん、ここでは厳密な再現性を要求しているのではない、再現したときに多少の変動があってもさしつかえないが、その変動性が法則性を利用するうえで支障のない範囲であることが必要である。この変動性を現実のデータからたしかめることはむずかしいことも多いが、例えば、§2に紹介する複合標本方式を利用して得られたデータの場合にはきわめて容易である。統計的方法の本質は、このような変動性を配慮した分析方法を与えることにあり、その点が数学的方法と根本的に異なるところである。

③について、目標母集団を悉皆調査して得たデータの場合を除いて、標本が母集団の一部である限り存在するのが標本誤差である。そして、標本が母集団からのランダムサンプリングにもとづくものであれば、推定値の標本誤差がデータから推定できることは衆知のことである。ただ、標本誤差を評価することの必要性は広く認識されているにもかかわらず、現実のサンプリング方式と推定方式に即した評価が、多くの場合実行されていないのは誠に遺憾である。単純ランダムサンプリングによる場合、あるいはそれが仮定できる場合には、標本誤差の評価は簡単であるため、その評価はなされていることも多く、また多くの統計的分析法がこのサンプリング法を前提としているので、その中で包摂的に処理されているケースも多い。単純ランダム以外の場合に、省略されることが多いのは、評価の公式が複雑であること、1調査あたりの項目数が多いこと、結果利用者が推定値のみで満足することが多いこと、等のためではないかと思われるが、これは困った問題である。単純ランダムの場合の公式で計算しておいて、経験的に称して、その2倍で済ませる、という乱暴なのを見かけることもしばしばである。仮説検定を行う場合には、標本誤差の評価如何によって、まったく逆の結論が導かれることもじゅうぶん起こり得る。その意味で、現実のサンプリング方式に忠実な評価式で、全項目でなくても、少なくとも主要項目については、計算を実行されることを強く望みたい。ただ、統計解析の手法の中には、単純ランダムサンプリング以外のサンプリング方式に対応した変動性を処理すること

が完成されていないものがある。しかし、複合標本方式によった標本の場合には、理論的にこのような制約はなく、また標本誤差の評価方式がきわめて簡便で、かつその簡便さがサンプリング方式・推定方式に依存しない、という大きい特徴を持っている。

最後に④について、一般的には、上掲の①②も非標本誤差に含めるべきであるが、それらを特掲しなかったことで、ここでは①②③以外の誤差を考えることにする。これに含まれるものとしては、無回答による誤差、調査対象の不理解・偽りによる誤差、調査員の犯すミス・ミスリーディングによる誤差などが含まれる。この範疇の誤差は複雑・多岐にわたり、統一的な処理法はないが、部分的には調査組織を確立整備することで、かなり軽減できるものである。無回答については、安易に回答者のみのデータ処理に走るのではなく、再三追求することも必要であり、とくに郵送調査の無回答処理は、面接法を併用する方法も考慮すべきである。また、別途データにもとづいて、回答者群と無回答者群の特性を比較検討することも有効である。調査対象に起因する誤差は、ある程度調査票の設計を注意深くし、調査組織を整備することで軽減できるが、調査対象の偽の回答は困ることが多い。これについては、著者の古い研究があるので、§3に紹介する。

2. 複合標本方式のすすめ

標本調査論の Text-book には、各種の抽出方法とともに、それぞれに対応した推定量の標本誤差の推定方式が克明に書かれているにもかかわらず、具体的な調査で標本誤差を忠実に計算することを省略しているケースが多い。これは、一つには計算公式が複雑で、計算機を利用するにもプログラム・パッケージがなく、個々にプログラミングをしなければならないことが一因ではないかと思われる。これに対し、Deming (1956) が提唱している複合標本方式 (replicated sampling plan) は、標本誤差の計算をきわめて単純化した点で、非常に有用な方法である。Deming のこの方法の端緒は、J. W. Tukey からの私信に記されたアイデアとして、すでに Deming (1950) に述べられているが、手法として Deming (1956) にまとめられ、Deming (1960) では全巻の大部分にわたって、多数の実例を通じて、くわしく解説されている。この方法が、標本誤差の推定のためにきわめて有用であるにもかかわらず、我が国ではあまり普及していない。それは多分標本誤差の推定に消極的な風潮があるためで

はないかと思われる。しかし、前述のように、データを分析し解釈するうえでの重要性を考え、ここで複合標本方式の簡単な解説をこころみる。

普通の標本誤差の推定は、得られた一組の標本をサンプリング方式に依存した適当な分類を行い、標本内の各種の内部変動を計算し、母分散・母集落間分散・母集落内分散等の母数を推定し、それらを抽出方式・推定方式に対応する分散式に代入する方法で実行される、それに対し複合標本方式では、最初から同じ抽出方式で複数組の標本を作り、それぞれから同じ推定方式で得た複数個の推定値間の変動を利用して標本誤差の推定を行うのである。最初から標本を複数組取るように計画さえしておけば、標本誤差の推定は抽出方式・推定方式には依存しないことが特徴で、複数組の標本を総合して一組と考えて推定値を求めれば、そのまま普通の意味での標本推定値になっているわけである。

今、同一の抽出方式で抽出された k 組の標本のそれぞれを第 i 副標本 ($i=1, 2, \dots, k$) と呼ぶ。第 i 副標本のサイズを n_i 、標本計量の合計を x_i 、標本平均を $\bar{x}_i$ とする。また k 組の副標本を総合して、 $x = \sum x_i$ 、 $n = \sum n_i$ と置けば、 $\bar{x} = x/n$ は標本全体での標本平均で、これは通常の意味での母平均の望ましい推定量と仮定しておく。各副標本は同一の抽出方式にもとづいているから、近似的に、

$$\bar{x} = \sum_{i=1}^k \bar{x}_i / k \quad (2.1)$$

と見てよい。すなわち、 $\bar{x}$ は k 個の $\bar{x}_i$ の標本平均であるから、その分散は、

$$V(\bar{x}_i) = \sigma^2(\bar{x}_i) / k \quad (2.2)$$

である。ここで $\sigma^2(\bar{x}_i)$ は k 個の $\bar{x}_i$ が標本値であるような母集団の母分散である。そしてこれは k 個の $\bar{x}_i$ から推定できる。したがって、

$$V(\bar{x}) \approx \frac{1}{k} \frac{1}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \quad (2.3)$$

さらに、 $\bar{x}_i$ はある程度の大きさ n_i の標本平均であることから、 $\bar{x}_i$ の母集団は近似的に正規分布に従うと見ることができ、 $k \leq 10$ 程度ならば、 $\bar{x}_i$ ($i=1, \dots, k$) のレンジ (範囲) $R(\bar{x}_i) = \bar{x}_{\max} - \bar{x}_{\min}$ を利用して $\sigma(\bar{x}) \approx R(\bar{x})/d_2$ なる推定も可能である。ここで、 d_2 は $N(0, 1)$ からの大きさ k の標本のレンジを R としたときの $E(R) = d_2$ である。一方、 $\sqrt{k} d_2$ を $k \leq 10$ の範囲の k について計算してみると、約 5% 程度の誤差の範囲で、 $\sqrt{k} d_2 \approx k$ とみなすことができ、結論として、

$$\sqrt{V(\bar{x})} \approx (\bar{x}_{\max} - \bar{x}_{\min}) / k \quad (2.4)$$

できわめて簡単に推定量 $\bar{x}$ の標準誤差の推定ができることが理解できよう。そしてこの推定は最初の抽出方式には無関係で、ランダムサンプリングでさえあれば、単純ランダムサンプリングでも多段サンプリングでも何でも成り立つ。

また、上記の $\bar{x}_i$ 、 $\bar{x}$ の代わりに比推定その他の任意の推定量 T_i 、 T に置き換えても、近似的に、

$$T \approx \sum_{i=1}^k T_i / k \quad (2.5)$$

が満足されるならば、同様に、

$$\sqrt{V(T)} \approx (T_{\max} - T_{\min}) / k \quad (2.6)$$

が利用できる。ここで、 T は偏りのある推定量であっても、 n_i が大で、偏りが無視できる程度のものならば、(2.5) の近似は成り立つものと見てよい。

さて、ここで k をいくらにするかの問題が残る。標本誤差の推定という立場だけからは、自由度が大きい方が望ましいから、 $k=10$ がもっともよいことになる。しかし、以下の事情からより小さい k が望ましいことも多い。

k 組の副標本を組織的に構成する方法として、Deming は (1956), (1960) でゾーン分け抽出法を提唱している。すなわち、1 段抽出法では抽出単位の抽出のさい、多段抽出法では第 1 段の抽出のさい、層化抽出法では各層内での抽出のさい、抽出単位のワク (リスト) を配列順に Z ずつの等間隔のゾーンに区切り、各ゾーンから k 個ずつの抽出単位を抽出して、抽出順に k 個の副標本に振り分ける方法である。この場合、ワク上の抽出単位の配列順に傾向性があったり、各層から少数個の抽出を予定しているときなどは、 k を大きくすることは、それだけ推定値の標本誤差を大きくする可能性がある。そのような懸念のある場合は、 $k=2$ とすることにし、そうでない場合には k をなるべく大きく、経験的には $k=6$ 程度でじゅうぶんのようである。ただ、 $k=2$ とするときは、(2.4) または (2.6) を利用することは、自由度が小さ過ぎ、よくない。Deming は (1956), (1960) で合併ゾーン (thick-zone) を構成して、別の公式で計算する方法を与えているが、(2.4) に比べるとかなり複雑なものになっている。その簡便化を図ったのが、著者 (1975) による「ランダム入れ替え」(random interchange) による方法である。

$k=2$ のゾーン分け抽出法では、各ゾーンで第 1 番目に抽出されたものを副標本 1、第 2 番目に抽出されたものを副標本 2 とする。そして、それぞれによる推定値を

表1 ランダム入れ換えの例示

ゾーン	原 標 本		ランダム 入 れ 換 え																					
	$(j = 0)$		$(j = 1)$				$(j = 2)$				$(j = 3)$				$(j = 4)$									
	副標本	副標本	入 れ 換 え	副標本	副標本	入 れ 換 え	副標本	副標本	入 れ 換 え	副標本	副標本	入 れ 換 え	副標本	副標本	入 れ 換 え	副標本	副標本							
	1	2		1	2		1	2		1	2		1	2		1	2							
	$n_1^{(0)}$	$x_1^{(0)}$	$n_2^{(0)}$	$x_2^{(0)}$	$n_1^{(1)}$	$x_1^{(1)}$	$n_2^{(1)}$	$x_2^{(1)}$	$n_1^{(2)}$	$x_1^{(2)}$	$n_2^{(2)}$	$x_2^{(2)}$	$n_1^{(3)}$	$x_1^{(3)}$	$n_2^{(3)}$	$x_2^{(3)}$	$n_1^{(4)}$	$x_1^{(4)}$	$n_2^{(4)}$	$x_2^{(4)}$				
①	②		③				④				⑤				⑥									
1	1	37	1	24	—	1	24	1	37	+	1	37	1	24	+	1	37	1	24					
2	1	18	1	15	+	1	18	1	15	—	1	15	1	18	+	1	18	1	15					
3	1	13	1	14	+	1	13	1	14	—	1	14	1	13	+	1	13	1	14					
4	1	9	1	11	—	1	11	1	9	—	1	11	1	9	+	1	9	1	11					
5	1	7	1	7	+	1	7	1	7	+	1	7	1	7	—	1	7	1	7					
6	1	6	1	6	—	1	6	1	6	+	1	6	1	6	—	1	6	1	6					
7	1	3	1	5	—	1	5	1	3	—	1	5	1	3	+	1	3	1	5					
8	1	2	1	2	—	1	2	1	2	+	1	2	1	2	—	1	2	1	2					
計	8	95	8	84	×	8	86	8	93	×	8	97	8	82	×	8	95	8	84	×	8	100	8	79
$\bar{x}_i^{(j)}$	11.88	10.50	10.75 11.63				12.13 10.25				11.88 10.50				12.50 9.88									
$R^{(j)}(\bar{x})$	1.38		0.88				1.88				1.38				2.62									

註 「入れ換え」欄の“+”は確率1/2で「入れ換えない」ことになったことを，“—”は「入れ換ええる」ことになったことを示す。

$\bar{x}_1^{(0)}, \bar{x}_2^{(0)}$ とし、そのレンジを $R^{(0)}(\bar{x}) (=|\bar{x}_1^{(0)} - \bar{x}_2^{(0)}|)$ とする。次に、ゾーンごとの副標本メンバーの入れ換えを 1/2 の確率で行う、すなわち、メンバーの入れ換えを行うか否かを 1/2 の確率で決めるのである。その結果として得られた副標本ごとの推定値およびそのレンジを、 $\bar{x}_1^{(1)}, \bar{x}_2^{(1)}, R^{(1)}(\bar{x}) = |\bar{x}_1^{(1)} - \bar{x}_2^{(1)}|$ とする。このランダム入れ換えを r 回繰り返し行って得た $R^{(j)}(\bar{x})$ ($j=1, \dots, r$) と最初の $R^{(0)}(\bar{x})$ とを利用して、

$$\sqrt{V(\bar{x})} \approx \frac{1}{2} \left(\frac{1}{r+1} \sum_{j=0}^r \{R^{(j)}(\bar{x})\}^2 \right)^{1/2} \quad (2.7)$$

で標準誤差の推定を行うのである。もちろん、 $R^{(j)}(\bar{x})$ は相互に強い相関はあるが、(2.7) の右辺は unbiased で、consistent であることが示され、4~5 回程度のランダム入れ換えの繰り返しで、 $R^{(0)}(\bar{x})$ のみによる場合よりも、かなり改善できることが経験的に判明している。表 1 は、 $k=2$ の原標本からランダム入れ換えを 4 回繰り返したところを例示したものである。

3. 非標本誤差の一考察

非標本誤差の処理法としては組織的な手法はなく、ただいくつかの特定の型のものについて多くの研究者が考察してきたが、ここで一つの場合として、「偽りの回答」の処理法を紹介する、これは著者の古い研究 (1963) であるが、この問題はアンケート調査でしばしば遭遇するもので、とくに有効であろう。

古くから統計調査論では、統計調査の妥当性をたしかめる意味で、本調査の終了直後、標本の一部を抽出して、同一項目について事後調査を行うことを奨めている。ただ、本調査と事後調査のあいだに不整合がある場合の明解な処理法がなかった。そこで、とくに yes-no 型の回答に限定して、2 調査間に不整合があった場合、調査対象が「嘘を答える確率」なるものを導入して、「嘘」がなかったと仮定したときの真の yes の割合の推定量を与えた。

表 2 本調査と事後調査との照合結果

本調査	事後調査		計
	yes	no	
yes	101	5	106
no	9	14	23
計	110	19	129

今、一例として、本調査の標本から大きさ $n=129$ の副次標本を選んで事後調査を行い、本調査の結果と照合したところ、表 2 のとおりの結果が得られた場合を考える。この表から、2 調査のいずれかに嘘を答えたものが 14 名いることがわかる。もっとも、2 調査とも yes または no と答えて、見掛け上は整合していても、両方とも嘘をついている可能性もある。このことをも考慮に入れて分析するのである。

表 3 照合表における記号

本調査	事後調査		計
	yes	no	
yes	n_{11}	n_{10}	$n_{1\cdot}$
no	n_{01}	n_{00}	$n_{0\cdot}$
計	$n_{\cdot 1}$	$n_{\cdot 0}$	n

表 2 を形式化して、各マス内の量をあらわす記号を表 3 に示すとおりとする。一方、

p : yes の真の割合

$q=1-p$: no の真の割合

p_t : 真を答える確率

$q_t=1-p_t$: 嘘を答える確率

とする。

ここで、 $n_{11}, n_{10}, n_{01}, n_{00}$ および $n(=n_{11}+n_{10}+n_{01}+n_{00})$ が与えられたときの、 p_t および p の最尤推定量は、

$$\hat{p}_t = \frac{1}{2} + \sqrt{\left(\frac{1}{2}\right)^2 - \frac{n_{10} + n_{01}}{2n}} \quad (3.1)$$

$$\hat{p} = \left(\frac{n_{11}}{n} - \hat{q}_t^2 \right) / (\hat{p}_t - \hat{q}_t), \quad [\hat{q}_t = 1 - \hat{p}_t] \quad (3.2)$$

で、これらの分散は、

$$V(\hat{p}_t) = \frac{1}{2n} \frac{(p p_t^2 + q q_t^2)(p q_t^2 + q p_t^2) p_t q_t}{(p-q)^2 p_t^2 q_t^2 + p q (p_t - q_t)^2 (p_t^2 + q_t^2)} \quad (3.3)$$

$$V(\hat{p}) = \frac{1}{n} \frac{(p p_t^2 + q q_t^2)(p q_t^2 + q p_t^2)}{(p_t - q_t)^2 (p_t^2 + q_t^2)}, \quad (3.4)$$

さらに分散の推定値は、

$$V(\hat{p}_t) \approx n_{11} n_{00} (n_{10} + n_{01}) A / [n \{n_{11} - n_{00}\}^2 \times (n_{10} + n_{01})^2 + \{nA - (n_{11} - n_{00})^2\} \times A(n_{11} + n_{00})] \quad (3.5)$$

$$V(\hat{p}) \approx n_{11} n_{00} / [nA(n_{11} + n_{00})] \quad (3.6)$$

ただし $A = 2(n_{11} + n_{00}) - n$

で与えられることが得られている。

表2のデータから p_i , p の推定値および分散の推定値を計算すると、

$$\hat{p}_i = 0.942, \quad V(\hat{p}_i) \approx (0.015)^2$$

$$\hat{p} = 0.881, \quad V(\hat{p}) \approx (0.031)^2$$

すなわち、嘘をつく確率が約6%あり、それを修正した yes の割合の推定値は88%となっている。

参考までに、嘘をつく確率による修正をまったく行わない yes の割合を求めると、

$$\text{本調査から} \quad \hat{p}_1 = \frac{106}{129} = 0.822$$

$$\text{事後調査から} \quad \hat{p}_2 = \frac{110}{129} = 0.853$$

で、上記の修正後の推定値は $\hat{p}_1$, $\hat{p}_2$ のいずれよりも大になっている。一般に、この処理法によると、 $p > 0.5$ ならば、無修正の場合よりも大きい値の $\hat{p}$ が、 $p < 0.5$ ならば、無修正の場合よりも小さい値の $\hat{p}$ が得られることになる。これは、「嘘」の存在を前提としているから当然のことである。

なお、本調査全体からすでに得られている推定値 $\hat{p}$ を、事後調査によって得られた嘘をつく確率 $\hat{p}_i$ で修正するには、

$$\hat{p}^* = (\hat{p} - q_i) / (\hat{p}_i - q_i) \quad (3.7)$$

でよいことも容易に導かれる。

この方法を実際の調査に適用した場合、調査員によるランダム・ミスが存在したとすれば、 q_i はそれをも含んだランダム・ミスの確率になっている。

あ と が き

上記は著者の専門分野である標本調査論の立場から、

行動科学関係の報告・論文を見聞したうえで感じた誤差についての印象をもとにして、書き綴ったものである。

ここで取り上げた内容は新しいものはほとんどなく、かつ衆知のことばかりであったが、調査の誤差についての認識を新たにして欲しいという願望から、あえて書くことにした。個々の報告・論文を誤差に関連して批評することはあえて避けたため、敘述がかなり抽象的になってしまった。「プログラミングが面倒なので、標本誤差の計算を省略した」という事態がなるべくなくなることを願って、著者の研究室でプログラム・パッケージの作成を計画していることを最後に付け加えておく。

参 考 文 献

- Asai, A. (1963). Some consideration on response errors of sample surveys. *J. of College of Arts & Science, Chiba Univ.* 4, 9-12.
- (1975). On a paired sampling plan. *J. of College of Arts & Science, Chiba Univ.* B-8, 53-64.
- Deming, W.E. (1950). *Some Theory of Sampling*. J. Wiley, New York.
- (1956). On simplification of sample design through replciation with equal probabilities and without stages, *J. of Amer. Stat. Ass.* 51, 24-53.
- (1960). *Sample Design in Business Research*. J. Wiley, New York. (斎藤・浅井・大沢訳：調査における標本設計，日科技連出版社)
- 斎藤金一郎・奥野忠一・浅井晃・他。(1979) 文部省認定社会通信教育「現代統計実務講座」(改訂版)，実務教育研究所。